

Interviewer Effects on Non-Differentiation and Straightlining in the European Social Survey

Geert Loosveldt¹ and Koen Beullens¹

In this article we examine the interviewer effects on different aspects of response styles, namely non-differentiation and straightlining, which in general refers to the tendency to provide the same answers to questions in a block of questions. According to research about response styles, the impact of the interviewer on this kind of response behavior is rare. Five blocks of items in the questionnaire in the sixth round of the European Social Survey (2012) are used in the analysis. These data also allow for an evaluation of the differences between countries in terms of non-differentiation and straightlining. Five different measurements of these aspects of response style are used in the analysis. To disentangle the impact of respondents and interviewers on these aspects of response style, a three-level random intercept model is specified. The results clearly show interviewer effects on the respondent's tendency to select a response category that is the same as the response category for the previous item. In some countries the proportion of explained variance due to differences between interviewers is larger than the proportion of variance explained by the differences between respondents.

Key words: Response style; three level random intercept model.

1. Introduction

Many survey questionnaires contain lists of statements about a particular topic or objects with the same response categories (for example, answers on a five-point scale ranging from 'strongly agree' to 'strongly disagree'). Respondents are asked to think about an object or statement and to select the response category that matches their opinion or position. Researchers assume that each respondent's rating is based on a thorough evaluation of all the response categories when answering this type of question. However, researchers are also aware that respondents sometimes insufficiently differentiate between questions and select the same response category for all the items or objects on a list, even though the items are not identical and may express something different. The tendency to provide the same answers to all of the questions in a block of questions about the same topic is called non-differentiation or straightlining. This kind of response style can be considered a source of systematic measurement error. It is clear that when all the items concerning a particular topic are formulated in one direction (positive or negative), straightlining can have a serious impact on the correlation between the items.

¹ Centre for Sociological Research, KU Leuven, Parkstraat 45, 3000 Leuven, Belgium. Emails: Geert.Loosveldt@kuleuven.be and koen.beullens@kuleuven.be

Straightlining can be considered as one of the types of respondent satisficing, a term coined by [Simon \(1956\)](#) and was later introduced in survey methodology research. Satisficing can occur when the cognitive demands of answering questions exceed the respondent's motivation and/or ability to optimize the response process ([Krosnick 1991](#)). Research into the stability of different response styles concludes that these are stable individual characteristics ([Weijters et al. 2010a](#)). This means that the respondent is mainly responsible for straightlining and that the respondent's characteristics and personality are relevant to explaining their response style. In their literature overview, [Van Vaerenbergh and Thomas \(2013\)](#) conclude that sociodemographic variables affect response styles, but the findings are not always consistent. There is more support for the relationship between personality and response styles, which is particularly true for extreme response styles. However, demographic and personality variables explain only a relatively small proportion of the variance in response styles, whereas culture and country-level characteristics seem to explain a relatively large proportion ([Van Vaerenbergh and Thomas 2013](#)). This result indicates that it is relevant and advisable to explain response styles not only based on respondent characteristics, but also on relevant contextual factors. In the current article, we consider the presence of an interviewer in face-to-face interviews as an obvious relevant contextual element. In the next section, we review literature and research concerning the impact of interviewers on response styles.

The importance of this article is twofold. First, we seek to provide evidence as to whether interviewers – as well as respondents – contribute to response styles involving non-differentiation and straightlining. Second, the European Social Survey (ESS) is used for the data analysis. This survey has become an increasingly important source of scientific output in the social sciences. According to [Malnar and Müller \(2014\)](#), using Google Scholar, some 89 publications based on ESS data could be found in 2003, and this number increased to 381 annual publications in 2013. As of June 2016, there are 94,317 registered data users of the ESS (url: http://www.europeansocialsurvey.org/about/user_statistics.html). For this reason, we seek to attract the attention of social science researchers in order to mitigate their scientific claims based on this face-to-face collected data source if interviewer variance is observed regarding response styles related to straightlining or non-differentiation.

2. Interviewers and Response Styles

Although the factor 'interviewer' is not completely absent in studies about response styles in face-to-face interviews, interviewers certainly do not have a dominant place in this type of research. [Hox et al. \(1991\)](#) identified interviewer effects on acquiescence after controlling for some relevant respondent variables. They measured acquiescence by the number of 'agree' responses to all the items on a balanced scale, and this indicates a respondent's systematic tendency to give answers irrespective of the content of the questions or items. In their analysis it was not possible to explain this interviewer effect by using the available interviewer characteristics. More recent research confirms the previously observed presence of interviewer effects on acquiescence, and shows that interviewer experience, after controlling for the length of the interviews, can explain a significant but small part of the variance (1.2–1.3%) in acquiescence across interviewers.

More experienced interviewers obtain higher levels of acquiescence than inexperienced interviewers, which means that interviewer experience is associated more clearly with a variance in acquiescence compared with the respondent's education level (Olson and Bilgen 2011). These results support the idea that, in face-to-face interviews, a response style is not just a matter of respondents' cognitive efforts, but also relates to how interviewers deal with this particular response behaviour. Olson and Bilgen (2011) conclude that models for acquiescence focus on respondent characteristics but tend to ignore the role of interviewers.

We examine whether the scarce results related to interviewer effects on acquiescence extend to the response style that can be termed non-differentiation, or straightlining. Based on the observation that a response style can be influenced by contextual factors, we expect that interviewers will also have an effect on a respondent's tendency to select the same or a nearby response category. We assume that in face-to-face interviews, not all interviewers react in the same way to this kind of response behaviour, and that some interviewers might be more inclined than others to facilitate or inhibit non-differentiation or straightlining. As a consequence, one can expect that differences between interviewers explain a significant proportion of the variability of such response styles. In the analysis of interviewer effects, we control for some respondent characteristics that are suggested to be related to respondent satisficing (motivation and ability). This also means that we take into account the differences between interviewers concerning the group of respondents. Based on the previous observed difference between countries in terms of response style, it also seems appropriate to evaluate the differences between countries.

All this makes clear that it is relevant to evaluate the impact of interviewers on straightlining and non-differentiation while controlling for some relevant respondent characteristics, and that it is advisable to do this for different countries. This general objective is specified in the data and models section. First, we start with a discussion of the different measurements of straightlining and non-differentiation, which will be the dependent variables in our analysis.

3. Measurement of Straightlining and Non-Differentiation

In relevant literature, several approaches are used to measure different types of straightlining. The starting point is the general definition of straightlining as the tendency to give the same answers to questions regardless of their content. Clear examples of straightlining can be observed in online surveys, where multiple items with the same response scale are displayed in a grid. In such a grid, the items are the rows and the response categories are the columns. A pure pattern of straightlining occurs when the selected answers are in a perfect vertical line, which means that the same response category was selected for each item. In their analysis of the association between speeding and straightlining in online surveys, Zhang and Conrad (2013) use eight grid questions with more than two statements and the number of grid questions on which respondents straightline (pure straightlining pattern) is used as a measurement of this response style. This specific operationalization of the measurement of straightlining can be used to discuss several aspects that are relevant to measuring straightlining.

3.1. *Different Aspects of the Measurement of Straightlining and Non-Differentiation*

3.1.1. Homogenous Versus Heterogeneous Sets of Items

In the example from [Zhang and Conrad \(2013\)](#), the topic of the items is the same within each grid (for example, (non)-working mothers, the role of father and mother in the household, etc.). This means that items in a grid are relatively homogeneous and express different aspects or nuances of the same topic. It is assumed that respondents take these nuances into account during the cognitive process in which they create and formulate their answers, and we therefore expect answers to be similar or consistent, but not identical. The substantially consistent answers are responsible for the correlation between the obtained answers for different items. It should also be noted that a response style such as non-differentiation or straightlining is an underlying factor that contributes to the correlation between items. When there is a pure straightline pattern for all respondents, the correlation between the items will be perfect. It is clear that this is not desirable.

Instead of using a homogeneous set of items about a topic, an alternative is to select a heterogeneous set of items, which can be presumed to be only moderately or poorly related. [Greenleaf \(1992\)](#) considers a set of items with low inter-item correlations as a prerequisite for creating a measurement of an extreme response style. For example, to create a set of heterogeneous items, [Weijters et al. \(2010b\)](#) randomly sampled 21 items from the same number of unrelated marketing scales. This procedure resulted in low inter-item correlations. Unrelated or poorly related items are considered as a necessary condition to ensure that the systematic tendency to select a response category is 'regardless of the content'. We have already noted that straightlining can increase the correlation between items. The question can be posed as to whether it is still possible to observe straightlining or non-differentiation with a set of independent items, because it appears difficult to observe the cause (response style) in a condition where the effect or result (correlation) is supposed to be completely absent.

3.1.2. General Straightlining or Non-Differentiation Versus Specific Response Categories

The second observation on the measurement procedure used in [Zhang and Conrad's](#) article (2013) is that straightlining is not specified for a response category. For example, no difference is made between the systematic use of extreme response categories or the middle scale category. Therefore, one can consider this a general measurement of straightlining in comparison with a specific measurement for the systematic tendency to select a specific response category, for example Extreme Response Style (ERS) or Midpoint Response Style (MRS).

3.1.3. Pure Pattern Versus Tendency

In the approach of [Zhang and Conrad \(2013\)](#), only a pure pattern – choosing the same response option for all the items in a grid – is considered to be evidence of straightlining. This is the traditional view on the operationalization of straightlining. In this article, we extend the operationalization of this concept in a more flexible way in order to obtain a

variety of tendencies of straightlining, or the closely adjacent concept of non-differentiation. Particularly, the number of items in a grid is not taken into account when measuring pure straightlining. This means that it is easier to fulfil the condition of straightlining when the grid contains only a small number of items (for example, three instead of eight). When the number of items in a block of questions on a topic increases, one can assume that it will be more difficult to observe a pure pattern. However, the absence of a pure response pattern does not mean that there is no clear tendency to systematically select a particular response category. The most obvious variant of the binary assessment (present or absent) of pure straightlining is counting the number of items with the same score (all rated the same, all but one rated the same, etc.) (Krosnick and Alwin 1988). One can also count the number of times a response category is selected and calculate the log odds ratio. The odds refer to the ratio of the number of times that a response category is selected, to the number of times that this category is not selected (Weijters et al. 2010). Other measurements of straightlining are based on the proportions or percentages of responses in a particular category.

3.1.4. Is the Order of the Responses Taken into Account?

This criterion is perhaps the crucial element that distinguishes straightlining from non-differentiation. The response sequence 7, 6, 6, 7, 6, 7 indicates the same degree of non-differentiation as the response sequence 6, 6, 6, 7, 7, 7. Nevertheless, the second sequence is more likely to provide evidence of straightlining. Non-differentiation is usually measured using a distance metric, such as the standard deviation of the responses, or the average square root of the absolute difference between any two answers from the same respondents to a block of questions (Chang and Krosnick 2009).

3.1.5. Response Scale Format

The eight grid questions used in the article by Zhang and Conrad (2013) do not all have the same response scale: there are six five-point scales ('fully disagree' to 'fully agree'), one different five-point scale ('certainly not' to 'certainly yes') and one three-point scale ('full-time', 'part-time', 'no job at all'). In their article, Weijters et al. (2010b) demonstrate that the labelling of the scale format components and the number of response categories affect different types of straightlining, and that accordingly, empirical results based on different scale formats may not be comparable. Although one might assume that the tendency to select the same response category will decrease when the number of response categories increases, the results do not support this assumption. This also means that respondents do not necessarily differentiate their answers when a scale is used that has more response categories. In fact, this is not what a researcher expects when deciding to use a response scale with more categories. Nevertheless, it seems necessary to evaluate straightlining or non-differentiation for a particular type of response scale.

3.2. Indicators of the Measurement of Straightlining and Non-Differentiation

The discussion of several characteristics that are relevant to qualify the measurement of straightlining or non-differentiation makes it clear that there is no evident simple and univocal measurement. Depending on the survey design characteristics (for example, the

mode and frequency of grid questions with a particular response scale) and the research questions, one measurement may emphasize a different aspect of the response style more than another.

Therefore, it seems more appropriate to opt for more than one measurement. This allows us to assess the sensitivity and robustness of the results for different operationalizations of the concept. Five different, but probably closely related, measurements are used in the analyses here.

3.2.1. Pure Straightlining and the Maximum Sequence of Identical Responses

In line with [Zhang and Conrad \(2013\)](#), straightlining is indicated by a 0-1 binary variable, where the presence of straightlining only applies if all the responses are identical. In the data used in our analyses (ESS), most blocks of questions are relatively long, so that pure straightlining is somewhat exceptional (<5%). In addition, because the analyses use a three-level (residual – respondent – interviewer) data structure, the multilevel models will be very likely to fail to converge. Instead, *the maximum string of identical responses* is determined for each respondent and for each block of questions. Item nonresponse (don't know, refusal, or no answer) breaks a sequence, even if the next response is identical to the previous. This indicator is labelled here as 'MAX'. For example, the maximum sequence in '7, 7, 7, 6, 6' is three; the maximum sequence in '7, 7, DK, 7, 7' is two. Notice that the absolute number is used. This measurement takes the order of response into account, but is not related to a particular response category.

3.2.2. The Percentage of Responses That are Identical to the Response to the Previous Question

In the response sequence 6, 6, 8, 8, 6, two out of four responses (50%) are the same as the previous ones (although there are five responses, the first evidently cannot be compared with a 'previous' one). In fact, one can consider this measurement as an indicator of *response inertia*. Similar to the previous measurement, the order or the sequence of the responses is important in order to assess straightlining. Unanswered questions can never contribute to the numerator determining the fraction, but always add to the denominator. For example, the sequence 5, DK, DK, 5 counts zero out of three potential straightline answers. This indicator is labelled here as '%STR'. The next two indicators do not take the order of the sequences into account.

3.2.3. The Standard Deviation of All the Responses of One Respondent in One Block of Questions

This indicates the degree to which respondents differentiate between questions. Higher scores indicate more differentiation (as opposed to the first two indicators, where higher scores indicate less differentiation or more straightlining). This indicator is labelled here as 'SD'.

3.2.4. Mulligans' Score

Mulligan's score is closely related to the standard deviations measurement for straightlining ([Chang and Krosnick 2009](#)). It is a distance metric, measuring the average

square root of the absolute difference between any two answers from the same respondents in a block of questions, or:

$$\binom{n}{2}^{-1} \sum_{q=1}^n \sum_{q'>q}^n \sqrt{|x_q - x_{q'}|}$$

where n is the number of questions in the grid, and x is the answer of the respondent to question q . Similar to other indicators, unanswered questions do not contribute to the calculation of the distance measurements. For example, the responses 4, 5, NA, 6 will generate the same score as if the sequence of scores was 4, 5, 6. This indicator is labelled here as 'MUL'.

3.2.5. The Average Distance Between Two Subsequent Answers

The last indicator combines the distance approach of indicators SD and MULL, and also takes the order of the responses into account. The average distance is determined between response q and the response to the previous ($q - 1$) question. For example the sequence 6, 5, 6, 4 will have a score of $(1 + 1 + 2)/3 = 1.33$, whereas the result for the sequence 6, 6, 6, 5 is $(0 + 0 + 1)/3 = 0.33$. This indicator is labelled here as 'DEV.PREV'.

It should be noted that none of these indicators will be capable of watertight detection of the response style that is intended to be measured. Although the response sequence 7, 7, 7, 6 is very likely to generate scores that indicate straightlining or non-differentiation, this sequence of responses can still be a truthful reflection of the respondent's beliefs or attitudes. For the purposes of this article, revealing false positives of this type is possible but not really problematic. The overall level of straightlining or non-differentiation is not of primary interest here, as long as it is equal among interviewers. Nevertheless, we expect to observe interviewer variance regarding these indicators, which in turn should alert researchers who use the ESS data (or other survey data that is prone to such interviewer effects) that the data is not faultless and that, as a consequence, it should be treated cautiously. As already mentioned, non-differentiation or straightlining may artificially increase the correlations between items and it is explicitly not expected that individual interviewers will advance such processes. Evaluation of these interviewer effects is the main objective of this article. Additionally, the presence of such interviewer effects should aid the data producers to invest more in interviewer (and questionnaire) management in order to avoid these unwanted effects.

4. Data and Models

Data from the European Social Survey Round 6 (ESS6) is eminently suitable for the analysis of interviewer effects in general and on straightlining in particular. The ESS6 was organized in 2012 in 29 European countries (see website: <http://www.europeansocialsurvey.org/data/download.html?r=6>) and the data allows us to evaluate interviewer effects on straightlining within and between countries. Five blocks of items in the questionnaire of the ESS6 (2012) are used in this analysis. A block consists of consecutive items measured on an eleven-point response scale. The topic of the items and the labels of the extreme points of the eleven-point response scale can vary within one

block. Therefore, a block is not necessarily a homogeneous set of items with the same eleven-point scale. Blocks 1, 4, and 5 are more homogeneous (the same topic and the same eleven-point response scale), whereas Block 2 and Block 3 are more heterogeneous (several topics and different eleven-point response scales). The number of the block corresponds to its order in the questionnaire.

- Block 1 (B2–B8). Political trust: seven items about trust in the police and several political institutions; one eleven-point scale (0 = no trust at all; 10 = complete trust).
- Block 2 (B18d–B25). Evaluation of politics and policy: nine items about the importance of and satisfaction with democracy and the state of the education and health services; five different eleven-point scales.
- Block 3 (D28–D35). Wellbeing: eight items about the time respondents have to do things they really want to do and how much of the time they generally are interested in, absorbed in or enthusiastic about what they are doing; four different eleven-point scales.
- Block 4 (E1–E15). Democracy in general: 15 items about democracy in general; one eleven-point response scale (0 = not at all important for democracy in general; 10 = extremely important for democracy in general).
- Block 5 (E17–E30). Democracy in the country: 14 items about democracy in the respondent's country (0 = does not apply at all; 10 = applies completely).

For each respondent, we can calculate each indicator presented in the previous section for each of the five blocks separately. Each of these indicators in each block can be considered to be a repeated measurement within a respondent producing this response style. Therefore, for each respondent, there are five measurements for each indicator (one for each block) nested within the respondent. This results in a three-level hierarchical data structure that can be analyzed using a three-level random coefficient model. The five measurements in each block are the first or lowest level (measurement level), the respondents are the second level (respondent level) and the interviewers are the third or highest level (interviewer level) in this hierarchical data structure. Within this structure, $INDIC_{bij}$ is the measurement of one of the five indicators in a block b for respondent i interviewed by interviewer j (with $b = 1, 2, 3, 4, 5; i = 1, \dots, I; j = 1, \dots, J$), therefore all five indicators are separately used as dependent variables in the specified models. The country level is not considered a level in the data structure, and the analysis is carried out separately for each country. The main reason for this choice is that countries might follow very different strategies to recruit, pay, train, and monitor their interviewers. The comparisons of the results of separate analysis in each country will clearly demonstrate the prevalence of the response style, as well as interviewer variances regarding the response style in each country.

The first model for $INDIC_{bij}$ in the three-level data structure is as follows:

$$INDIC_{bij} = \gamma_{000} + \sum_{b=2}^5 \gamma_{b00} Block_b + \mu_{0j} + k_{0ij} + \varepsilon_{bij} \quad (\text{Model 1})$$

In this model, the only independent variable is the block information for which 5–1 parameters (the first block is the reference category) are accommodated, and μ_{0j} , k_{0ij} , and ε_{bij} are respectively the unique parts of the intercepts at the interviewer level, the

respondent level and the error at the measurement level in the block. The variances of these unique parts of the intercept are, respectively, $\sigma_{\mu 0}^2$, σ_{k0}^2 , and σ_{ε}^2 . With this model, we can break down the variance across the three levels and calculate the proportion of variance explained by the respondent ($\rho_{\text{respondent}}$) and the interviewer ($\rho_{\text{interviewer}}$) (Hox 2010; Loosveldt and Beullens 2013). The expressions for the proportions of explained variance are:

$$\rho_{\text{respondent}} = \frac{\sigma_{k0}^2}{\sigma_{\mu 0}^2 + \sigma_{k0}^2 + \sigma_{\varepsilon}^2};$$

$$\rho_{\text{interviewer}} = \frac{\sigma_{\mu 0}^2}{\sigma_{\mu 0}^2 + \sigma_{k0}^2 + \sigma_{\varepsilon}^2};$$

Both expressions are appropriate to evaluate the impact on the indicator of the response style both of interviewers and respondents. It is preferable for the proportion of variance explained by the interviewer to be small, or at least much smaller than the proportion of variance explained by the respondent.

The interpenetration of interviewers and areas was not accommodated during the design stage. This means that area effects may erroneously be taken for interviewer effects (and vice versa). Therefore, we extend the model by adding covariates at the respondent level in order to make the groups assigned to the interviewer more similar. A first extension (Model 2) includes the following variables:

- level of education of the respondent (a seven-point EISCED scale),
- gender of the respondent,
- age of the respondent, and
- rank (logarithm). This count variable indicates the chronological rank of the respondent within each interviewer. It may be expected that a response style is more likely to be observed as the interviewer becomes more familiar with the survey or its questionnaire.

After including these covariates, the model is estimated as if all respondents have a similar age, gender, level of education, and rank within the interviewer. In Model 3, two area variables are included:

- Density: a five-point scale indication of population density, self-reported by the respondent.
- Region: a geographical area that usually coincides with provinces, counties or any other NUTS2 or NUTS3 subnational entity.

Adding these variables may have a strong filtering effect, distinguishing area and interviewer effects. However, since interviewers are recruited, trained, monitored, and coached along the same geographical lines, real interviewer effects may erroneously be taken for area effects. In this way, a larger than ideal amount of real interviewer effects may be separated out of the interviewer variance. A final extension (Model 4) is made by adding the variables that indicate the motivation and cognitive skills of the respondent, as assessed by the interviewers:

- RESBAB: the respondent tried to answer the questions to the best of his/her ability (five-point scale).
- RESUNDQ: the respondent understood the questions (five-point scale).

A potential problem of adding these covariates to the model is that interviewers rate their own respondents. Because it is very likely that these variables are prone to containing interviewer effects themselves (on average countries show intra-interviewer correlations of 0.34 for RESBAB and 0.20 for RESUNDQ), there is a risk that these variables artificially explain too much interviewer variance regarding the response style of non-differentiation or straightlining.

We run the models per country. This is particularly relevant, because the main responsibilities for interviewer management are located at the country level. Each of the participating countries in the ESS needs to recruit, train, remunerate, and monitor its own interviewers. From the ESS6 documentation report (which can also be accessed using the link shown earlier), it becomes immediately clear that countries do not follow a uniform approach regarding the length of interviewer training, the materials they use for the training, interviewer payment, and so forth. Therefore, providing results per country seems to be most appropriate.

5. Results

We start with descriptive results for one of the indicators of non-differentiation and straightlining in the ESS6 (%STR). The first column in [Table 1](#) (Frequency) shows the frequencies with which the categories on the eleven-point scale are chosen. In total, 2,897,669 answers are considered, originating from 54,673 respondents in 29 countries, with each respondent giving 53 answers on an eleven-point scale ($53 \times 54,673 = 2,897,669$). Each of the 53 items belongs to one of the five blocks.

Table 1. Frequencies and percentages of straightlining on an eleven-point scale for 53 items and 54,673 respondents (ESS6).

Scale point	Frequency %	%STR %
0	5.07	48.91
1	2.67	29.84
2	4.08	23.97
3	5.37	22.28
4	5.60	20.45
5	11.26	28.53
6	8.41	21.70
7	11.47	26.08
8	13.69	31.58
9	8.96	33.49
10	19.42	65.53
Refusal (77)	0.16	61.11
Don't know (88)	3.78	43.35
No answer (99)	0.08	51.06

It is apparent that scale-point 10 is chosen most frequently, followed by 8, 7, and 5. Answer 1 and the nonsubstantive answers (don't know, refusal, and no answer) are given the least often. The second column in Table 1 (% STR) shows which scale points are more prone to having the same answer as that given to the preceding question (the second indicator as presented in Section 3: %STR). For example, 48.91% of all the 0 answers were also 0 answers to the previous item. From Table 1, it seems that straightlining or non-differentiation is more likely to occur in the extreme categories, as well as for nonsubstantive categories. However, we do not consider nonsubstantive answers to count as non-differentiation or straightlining in our analysis. It should be noted that this second column is not based on 53 items, but on 48 items (53–5), because providing a similar answer as that to the previous question cannot be assessed for the first item of each block.

Figure 1 shows the average percentage of straightline answers (%STR) in each country for the five blocks and for all the blocks together. The means in the subtitle of the figure are the mean percentages of straightline answers (%STR) in the block for all respondents. For example, for the first block on political trust containing seven items, on average 34.95% of

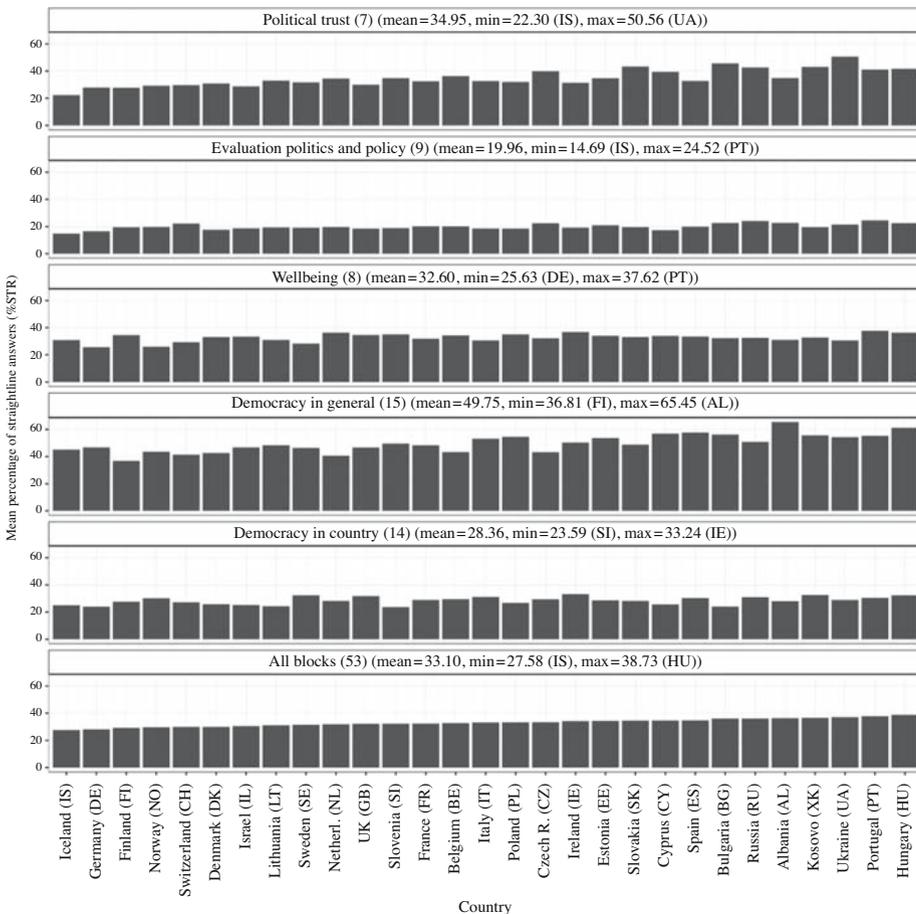


Fig. 1. Mean percentage of straightline answers (%STR) in separate blocks and all blocks together in each country (ESS6).

the answers are identical to the answer to the previous question, with Iceland (IS) showing the minimum (22.30%) and Ukraine (UA) the maximum at 50.56%. The overall view makes clear that there are differences in straightlining between blocks and within blocks between countries. The average percentages of straightlining are especially high for the questions about democracy in general (Block 4). In this block, the mean percentages of answers that are the same as the answers to the previous items are about 60 per cent in some countries. The average percentages are much smaller for the similar questions about democracy in the respondents' own country (Block 5). The more heterogeneous Block 2, with questions about the evaluation of politics and policy, seems to be less sensitive to the tendency to select a response category that is the same as that for the previous item. This makes clear that the measurement used is indeed, as expected, also affected by the homogeneity of the items used. On the other hand, the other block, which was characterized as more heterogeneous with questions about wellbeing, shows higher percentages of straightlining, and is comparable with the other more homogeneous blocks.

So, Figure 1 presents the information concerning one of the five indicators of straightlining and non-differentiation (%STR). Similar figures for the four remaining indicators ('MAX', 'SD', 'MUL', 'DEV.PREV') are available on online supplemental file of Journal of Official Statistics website (available at: <http://dx.doi.org/10.1515/jos-2017-0020>). In our analysis, we also used the four other indicators, namely 'Max', 'SD', 'MUL', and 'DEV.LAST'.

For each respondent and each of the five blocks, each of the five indicators can be determined based on the sequence of the answers given. Table 2 shows the average correlations between these five indicators, only considering measurements within the same block. This means that, for example, 0.89 is the average of five correlations: the correlation between '%STR' and 'MAX' in Blocks 1, 2, 3, 4, and 5. These correlations may be quite high because the indicators are all measured based on the same sequences of answers. Therefore, Table 3 shows the same average correlations, but only correlations between different blocks are allowed to contribute. For example, 0.18 is the average of the correlation between %STR and MAX in ten combinations of Blocks (Block 1 and Block 2, Block 1 and Block 3, . . . Block 4, and Block 5). These average correlations tend to be much lower. It is apparent that two groups of indicators can be distinguished: '%STR' and 'MAX' (similarity measures) tend to be similar, and 'SD', 'MUL' and 'DEV.PREV' (distance measures) also tend to cluster.

The key research question in this article is how much of the observed variability in these indicators of straightlining or non-differentiation can be attributed to the respondent level

Table 2. Average correlations between the five indicators measured within the same blocks. ESS6, 54,673 respondents.

	%STR	MAX	SD	MUL	DEV.PREV
%STR	1.00				
MAX	0.89	1.00			
SD	-0.32	-0.30	1.00		
MUL	-0.60	-0.59	0.87	1.00	
DEV.PREV	-0.55	-0.48	0.83	0.80	1.00

Table 3. Average correlations between the five indicators measured within different blocks. ESS6, 54,673 respondents.

	%STR	MAX	SD	MUL	DEV.PREV
%STR	0.19				
MAX	0.18	0.17			
SD	-0.01	0.01	0.21		
MUL	-0.08	-0.07	0.13	0.12	
DEV.PREV	-0.07	-0.05	0.17	0.14	0.17

and in particular, how much to the interviewer level. We assume that a response style is not only a matter of respondent behaviour, but in line with the results of [Olson and Bilgen \(2011\)](#), that it may be affected by the impact of the interviewer. As a result, we expect to observe that some part of the variance of any of the five indicators for non-differentiation or straightlining as a response style is explained by the interviewers. Because the five indicators tend to be correlated, but potentially measure different aspects of the response style related to straightlining or non-differentiation, it seems appropriate to present the results for all five indicators. In this way, we can assess to what extent our analysis of interviewer and respondent variance is sensitive to the choice of indicator. First, we will discuss the results regarding one indicator ('%STR'), after which a summary of the four remaining indicators will be provided.

[Figure 2](#) shows per country how much variance of '%STR' can be attributed to the interviewer level (black lines) and to the respondent level (grey lines). These lines result from applying Model 1 and its extension when respondent background characteristics are added (Model 2), subsequently area variables are added (Model 3), and variables are added that indicate the motivation and cognitive skills of the respondents (Model 4). The variance components under Model 1 are shown on the left sides of each subgraph and connected with the variance components of the three other models. In some countries, the estimated interviewer effect depends on the applied model. This can be observed in Bulgaria, Hungary, Italy, Kosovo, Lithuania, Portugal, and Slovenia. Particularly between Model 2 and 3 (where area variables are added), the shift is strongest. Nevertheless, as already mentioned, this covariate information may not only contain regional effects between respondents (which is the major reason for including these variables), but may also provide unintended information explaining differences between interviewers, as their interviewing styles may also be trained locally. In the other countries, there are no considerable changes in the estimated shares of variances of interviewers and respondents according to the applied model. Only the estimated impact of the respondents may in some countries differ between Model 1 and 2 (after the inclusion of the respondent variables age, gender, and level of education, and the variable indicating which chronological rank the respondent has within the interviewer). Such clear effects can be observed in Denmark and Belgium.

There are many countries in which the interviewers tend to have a rather small share in the variance of the indicator '%STR'. These countries are Albania, Belgium, Denmark, Finland, France, Germany, Iceland, Israel, Italy, the Netherlands, Norway, Spain, Switzerland, and the United Kingdom. In these countries, it can be observed that the black

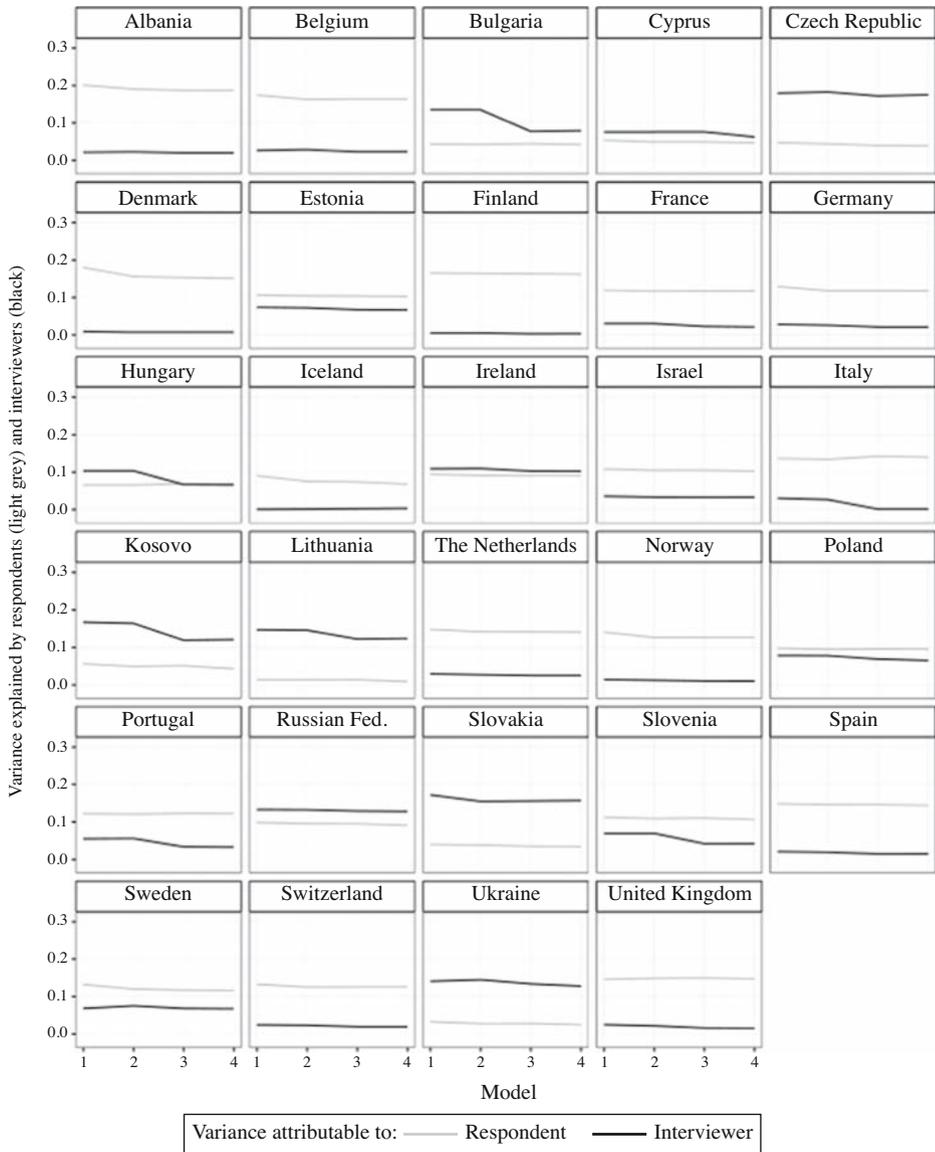


Fig. 2. Intra-respondent (light grey) and intra-interviewer (black) correlations for the ‘%STR’ indicators for four model specifications in 29 countries of the ESS6.

line is very close to the origin (although in most cases it is still statistically significant from zero). In some other countries, interviewer effects on ‘%STR’ are more substantial, although the respondent effects on this indicator are also considerable. This can be observed in Cyprus, Estonia, Hungary, Ireland, Kosovo, Poland, Portugal, Russian Federation, Slovenia, and Sweden. Finally, in a few countries, the interviewer effects clearly surpass the impact of the respondents. This applies to Bulgaria, the Czech Republic, Lithuania, Slovakia, and Ukraine. It is clear that in these countries, intra-interviewer correlations of 0.10 and more are no exception. Therefore, in these countries,

interviewers explain a considerable amount of the variation in the response style '%STR' and it becomes difficult to claim that straightlining or non-differentiation is a trait that is only attributable to respondents.

The graphs of the four remaining indicators ('MAX', 'SD', 'MUL', 'DEV.PREV') are available on online supplemental file of Journal of Official Statistics available at: <http://dx.doi.org/10.1515/jos-2017-0020>. Most of the results as portrayed in Figure 2 also apply to the other indicators of straightlining and non-differentiation. Nevertheless, some noteworthy differences can be summarized as follows. The indicators 'MAX' and 'MUL' tend to show lower levels of variance components for both interviewer and respondents. 'SD' tends to show higher respondent effects as compared with the four other indicators. Generally, the estimates of interviewer variance are more stable across indicators, whereas the estimates of respondent variance are more varied across indicators. For some countries, some noteworthy remarks need to be made. For Bulgaria, in Figure 2 the respondent variance is about 0.05 for '%STR'. For the other four indicators, this variance reduces to <0.01 . This also applies to Kosovo. In Israel, the interviewer estimate for '%STR' is about 0.04 and is therefore much smaller than the respondent impact (about 0.1). However, in the case of 'DEV.PREV' and 'SD', both interviewer and respondent variance are estimated at about 0.1. In Portugal, interviewer variance is estimated to be smaller than the respondent variance regarding '%STR'. However, regarding the four other indicators, respondent and interviewer variance have about the same magnitude.

The parameter estimates for all the added covariates of Model 4 can be found in the online appendix. However, there do not seem to be clear patterns as to which variables are (strongly) predictive of the response style. Gender does not have a clear effect on non-differentiation or straightlining, although in some countries, men tend to have higher scores for 'SD', 'MUL', and 'DEV.PREV'. Generally, age is positively related to '%STR' and 'MAX'. Strangely, age also positively relates to larger differences between the answers (as indicated by 'SD', 'MUL', and 'DEV.PREV'). In some countries, however, the age effect is in the opposite direction. Level of education seems to be related to the response style in some countries, but the direction of the relationship is not very clear, as opposing effects are observed. Population density does not seem to be closely related to the response style, and in the few countries in which effects are observed, they are inconsistent. Because the list of the different provinces or counties used in Model 3 and Model 4 is very long, the related parameter estimates are not provided in the appendix. Nevertheless, regions can differ considerably from one another. This is particularly so in Bulgaria, Hungary, Italy, Lithuania, Portugal, Slovenia, Slovakia, and Kosovo. In these countries, there is a noticeable decrease of the intra-interviewer correlation between Model 2 and Model 3, where the latter includes regional covariates and the former does not.

The effect of the chronological rank (logarithm) of the respondent within the interviewers is not very strong, but it is consistent. Respondents who are interviewed later by the same interviewer tend to show higher levels for '%STR' and 'MAX', indicative of straightlining, and lower levels for 'SD', 'MUL', and 'DEV.PREV', indicative of non-differentiation. In only a few countries, more motivation on behalf of the respondent – as observed by the interviewers (RESBAB) – tends to be related to more differentiation. Whether the respondent understood the questions (as assessed by the interviewer,

RESUNDQ) tends to relate somewhat more strongly to the five indicators, although the directions of these relationships are rather ambiguous. Respondents showing that they understood the questions tend to straightline more ('%STR' and 'MAX'), but also tend to differentiate more ('SD', 'MUL', and 'DEV.PREV').

6. Conclusion and Discussion

In an article about the past, present, and future of total survey error, [Groves and Lyberg \(2010\)](#) conclude that the study of the interplay of various different error sources must be part of the agenda for future survey methodological research. In line with this conclusion, in the current article we try to combine research into response styles with the assessment of interviewer effects, a combination that is rare in survey methodological research. Research concerning response behaviour makes clear that the process can sometimes be characterized by a response style showing a lack of effort to obtain adequate and correct answers. In the total survey error framework, response styles are a source of measurement error for which the respondent is responsible. However, research into interviewer effects makes clear that interviewers can have an impact on the registered responses, and that interviewers can also be considered as another source of measurement error. In the current article, we evaluate the impact of interviewers on the respondent's response style: particularly the tendency to provide the same answer as the one to the previous question (straightlining) and non-differentiation. It should be emphasized that we do not seek to provide evidence that respondents are straightlining or failing to adequately differentiate between questions, but instead we want to provide evidence of the extent to which interviewers mediate these processes. Using different indicators and different data sets from various countries, our results clearly illustrate that in most countries, interviewers have a significant impact on these response tendencies. This makes it clear that analyzing the interplay between the respondent and the interviewer as sources of measurement error was fruitful and that it allows for an interpretation of response styles from a different perspective. Response style is not only a matter of the respondent's cognitive processes, motivation or other characteristics. This interpretation is too limited. Response styles are also influenced by situational factors, and in face-to-face interviews the interviewer is not a negligible factor. The assessment of interviewer effects is a diagnostic analysis. Our results clearly indicate that interviewers can have an impact on response patterns, but we do not know how and why. It is possible that interviewers are suggestive, or that they reinforce a respondent's tendency to select similar response categories. A straightforward way to find out how response styles operate during an interview would be to record and analyze the interviewer-respondent interaction. The results of this type of analysis can also be used during training to remedy interviewers' shortcomings. The observation of substantial interviewer effects highlights the importance of training, and the results of this study reinforce that. During interviewer training, it is advisable to pay adequate attention to response styles. It is necessary to ensure both that interviewers do not induce a response style, and that they know how to handle different kinds of response behaviour during the interview.

Currently, there is an increased attention by survey researchers and practitioners to collect paradata, also during data collection. One of the aims is to monitor and potentially

improve the quality of the data while the data collection process is ongoing. In that sense, regular monitoring of the data of completed interviews can be done in order to assess the degree to which certain interviewers tend to show signs of unfavourable patterns in the obtained answers from their respondents. Such close quality control might be the basis for continual interviewer coaching during the fieldwork.

The observed differences between countries are remarkable and the interpretation of this in terms of ‘cultural differences’ seems too general and inconclusive. Although it is possible that in some countries straightline answers or non-differentiation are in agreement with the ‘true attitude’ of the respondent (for example, a very negative evaluation of all aspects of the democratic system), this cannot be an explanation for the differences between interviewers. Differences in ‘survey culture and practices’ and in fieldwork capacity are probably responsible for differences between countries. The fact that in some countries the differences between the interviewers explain more variability in response tendency than the differences between respondents must, at least, be considered an urgent call to closely monitor the way fieldwork procedures (interviewer training and briefings, follow up of the interviewers during the fieldwork, feedback for interviewers, etc.) are implemented in these countries. One must be aware that these differences in interviewer effects on response style can influence substantive comparison across countries, for example because straightlining or non-differentiation may artificially inflate correlations between survey items. It is necessary to ensure that differences between countries in ‘survey culture and practices’ are not interpreted as real cultural differences.

7. References

- Chang, L. and J. Krosnick. 2009. “National Surveys via RDD Telephone Interviewing versus the Internet. Comparing Sample Representativeness and Response Quality.” *Public Opinion Quarterly* 74: 641–678. Doi: <http://dx.doi.org/10.1093/poq/nfp075>.
- Greenleaf, E. 1992. “Measuring Extreme Response Styles.” *The Public Opinion Quarterly* 56(3): 328–351. Doi: <https://doi.org/10.1086/269326>.
- Groves, R. and L. Lyberg. 2010. “Total Survey Error: Past, Present, and Future.” *The Public Opinion Quarterly* 74(5): 849–879. Doi: <http://dx.doi.org/10.1093/poq/nfq065>.
- Hox, J. 2010. *Multilevel Analysis: Techniques and Applications*. 2nd ed. New York: Routledge.
- Hox, J., E. de Leeuw, and I. Kreft. 1991. “The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: a Multilevel Model.” In *Measurement Errors in Surveys*, edited by P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman. New York: Wiley.
- Krosnick, J. 1991. “Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys.” *Applied Cognitive Psychology* 5(3): 213–236. Doi: <http://dx.doi.org/10.1002/acp.2350050305>.
- Krosnick, J. and F. Alwin. 1988. “A Test of the Form-Resistant Correlation Hypothesis: Ratings, Rankings, and the Measurement of Values.” *Public Opinion Quarterly* 52(4): 526–538.

- Loosveldt, G. and K. Beullens. 2013. "The Impact of Respondents and Interviewers on Interview Speed in Face-to-Face Interviews." *Social Science Research* 42(6): 1422–1430. Doi: <http://dx.doi.org/10.1016/j.ssresearch.2013.06.005>.
- Malnar, B. and K.H. Müller. 2014. *Surveys and Self-Reflexivity: A Second-Order Study of the European Social Survey (ESS)*. Wien: Echoraum.
- Olson, K. and I. Bilgen. 2011. "The Role of Interviewer Experience on Acquiescence." *Public Opinion Quarterly* 75(1): 99–114. Doi: <https://doi.org/10.1093/poq/nfq067>.
- Simon, H. 1956. "Rational Choice and the Structure of the Environment." *Psychological Review* 63(2): 129–138. Doi: <http://dx.doi.org/10.1037/h0042769>.
- Van Vaerenberg, Y. and T. Thomas. 2013. "Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies." *International Journal of Public Opinion Research* 25(2): 195–217. Doi: <https://doi.org/10.1093/ijpor/eds021>.
- Weijters, B., E. Cabooter, and N. Schillewaert. 2010b. "The Effect of Rating Scale Format on Response Styles. The Number of Response Categories and Response Category Labels." *International Journal of Research in Marketing* 27(3): 236–247. Doi: <http://dx.doi.org/10.1016/j.ijresmar.2010.02.004>.
- Weijters, B., M. Geuens, and N. Schillewaert. 2010a. "The Stability of Individual Response Styles." *Psychological Methods* 15(1): 96–110. Doi: <http://dx.doi.org/10.1037/a0018721>.
- Zhang, C. and F. Conrad. 2013. "Speeding in Web Surveys: The Tendency to Answer Very Fast and its Association With Straightlining." *Survey Research Methods* 8(2): 127–135. Doi: <http://dx.doi.org/10.18148/srm/2014.v8i2.5453>.

Received January 2016

Revised February 2017

Accepted March 2017